Contribution ID: **50** Type: **not specified**

Towards better decoding and language model integration in sequence to sequence models

The recently proposed Sequence-to-Sequence (seq2seq) framework advocates replacing complex data processing pipelines, such as an entire automatic speech recognition system, with a single neural network trained in an end-to-end fashion. In this contribution, we analyse an attention-based seq2seq speech recognition system that directly transcribes recordings into characters. We observe two shortcomings: overconfidence in its predictions and a tendency to produce incomplete transcriptions when language models are used. We propose practical solutions to both problems achieving competitive speaker independent word error rates on the Wall Street Journal dataset: without separate language models we reach 10.6% WER, while together with a trigram language model, we reach 6.7% WER.

Author: NAVDEEP , Jaitly (Google Brain)

Presenter: NAVDEEP , Jaitly (Google Brain)